面向政府决策需求的社会舆情信息语义组织研究*

王曰芬[1,2] 邢梦婷[1]
[1](南京理工大学经济管理学院 南京 210094)
[2](江苏省社会公共安全科技协同创新中心 南京 210094)

摘要:【目的】对海量异构的社会舆情信息进行语义组织，揭示社会舆情信息之间的多维度关联，以为后期的挖掘、分析和决策支持提供重要支撑。【方法】从政府决策需求出发，建立基于本体的社会舆情信息语义组织框架，根据社会舆情信息本体的特点，采用七步法的本体构建方法，以 ABC 事件本体为参考模型，明确其中的概念以及概念间的关系，建立社会舆情信息之间的语义关联。【结果】构建适用于社会舆情信息描述的 POIOM 模型，完成本体管理模块的设计。【局限】论文以访谈和统计分析为主要需求获取渠道，有待更多的政府决策者用户群体对需求加以验证和完善。【结论】POIOM 模型对社会舆情信息具有良好的描述能力，从内容角度深层次地揭示同一案例不同特征项内容之间的关联，以及多舆情事件之间的逻辑关系和内在联系，基本验证了基于需求的社会舆情信息语义组织方法的可行性和有效性。

关键词: 政府决策需求 社会舆情信息 本体 语义组织
分类号: G350

1 引 言

随着新技术与新媒体的发展，社会舆情信息也凸显出海量(Volume)、瞬刻(Velocity)、全息(Variety)、价值低密(Value)、虚实莫辨(Veracity)的大数据特征，同时伴随着各种舆情事件的风起云涌，社会舆情的测度和分析日益复杂化，急需一种有效的社会舆情信息组织方式将多种媒介的各类社会舆情信息融合起来，实现社会舆情信息的细粒度揭示和多维度关联，以为后期的挖掘、分析和决策支持提供重要支撑。

近年来，国内外都十分关注网络环境下信息组织系统的理论构建、技术及实践研究，诸如概念关系揭示、语义网络构建、知识的语义化表达、叙词表、语义网络、主题图、概念图及本体等信息组织技术及其实践应用研究等[1]。已有一些学者和科研单位也专注于知识组织工具的研究，并有成功的实践案例。国外的 On-To-Knowledge[2]、KAON[3]、统一医学语言系统(UMLS)[4]和国内的中国医学语言系统[5]、中国农业知识本体[6]都是具有代表性的基于语义的网络信息组织的实践应用成果，国内的代表性研究还有董慧等[7-8]和申欣等[9]的研究成果，分别为数字图书馆资源组织和语义信息检索做出了一定的贡献，当前国内存在的问题，主要是面向具体领域的实际应用项目太少。

基于语义的舆情信息组织方面的技术主要涉及本体、语义网等，主要的研究成果如: Kinsella 等[10]利用语义网技术揭示了社会网络中用户及其生成内容之间的关系; Sleeman 等[11]提出了基于 FOAF 的社会语义网络研究方法，采用规则和机器学习的相关技术发现 FOAF 文档，并提取网络参与者的相关信息; 王兰成等[12]提出一种基于语义相似度的 Web 文本分类方法;

通讯作者: 王曰芬, ORCID: 0000-0002-7143-7766, E-mail: yuefen163@163.com。

*本文系国家社会科学基金重点项目“大数据环境下社会舆情与决策支持方法体系研究”(项目编号:14AZD084)、江苏高校优势学科建设工程资助项目和江苏高校哲学社会科学重点研究基地“社会计算与舆情分析”(培育点)的研究成果之一。

随后又提出基于情报语义关联集成分析的新方法ISAIA，并运用在智能网络舆情的热点发现与倾向分析研究中[13]。现有舆情本体方面的研究，主要是采用内容分析的方法，借助人工编码或专门的分析软件解读舆情本体中高频词汇所反映的现象，且这些研究主要是以典型案例为研究对象，如凤凰古城收费事件[14]、立顿茶包事件[15]等。可以发现基于语义的舆情信息组织的相关研究受到越来越多学者的关注，但未形成一定的规范和标准；此外，以典型案例为依托的研究，虽然问题说明更明确具体，但广泛适用性和可扩展性较弱，还需进一步的深入研究。

本研究借助于本体的思想与技术，将零散的社会舆情信息基于语义关联视角进行组织，以经典的ABC事件本体[16]为参考，建立社会舆情信息本体模型POIOM，基于信息宽时间、广维度和多事件的多维度视角，从内容角度深层次地揭示同一案例不同特征项内容之间的关联，以及多舆情事件之间的逻辑关系和内在联系，将多种媒介的各类社会舆情信息融合起来，有利于揭示舆情事件相关信息深层次的含义，从而对政府应对措施的制定提供更好的决策支持。

2　面向政府决策需求的社会舆情信息语义组织的意义与思路

2.1　面向政府决策需求特征的社会舆情信息特征分析

通过调研了解到：政府决策过程的不同阶段对信息的需求也存在差异，其主要可以分为来源、载体等外部特征和包含涉及群体、时间点、地域、敏感话题等内容特征的需求。本研究结合社会舆情信息的基本特征(包括事件类型、参与主体、主题、时间、地点等)，参考《网络舆情蓝皮书》[17]、《中国社会舆情与危机管理报告》[18]、《危如朝露：2010~2011中国网络舆情报告》[19]等舆情报告，政府决策需求特征可从数据来源、数据载体形式、事件类型、涉及群体、时间点、地域、敏感话题、热门话题这几个方面进行提炼，结果具体如表1所示。

以近几年的热点社会舆情案例为对象，具体选取2007年至2014年时间段所发生的舆情事件案例，来源于各大舆情报告以及互联网上报道发布的各大热点社会舆情事件，如人民网每年度发布的《中国互联网

表1　政府决策需求

维度	政府关注的舆情事件特征
数据来源	主流的新闻类网站、有影响力的BBS论坛和微博客等
数据载体形式	以文本信息为主，辅以图像、音频、视频等形式
事件类型	社会民生、反腐倡廉、司法执法、时事外交、事故灾难等
涉及群体	公务员、政府官员、医生、城管、学生、老人、弱势群体等
时间点	重要时间点周围的时间段发生的舆情事件，比如两会、国庆等
地域	北京、广东、湖北、江苏、重庆、湖南等舆情事件频发地域，台湾、新疆等敏感地域，全国范围性
敏感话题	涉及台独、恐怖分子等敏感词汇的话题
热门话题	居于新闻排行榜、评论排行榜中，跟帖点击频率高、微博博客转发次数多的话题

舆情分析报告》等，并依据上文分析得到的政府决策者在群体、时间点、地域、敏感话题等内容特征方面的需求为依据，以人工筛选的方式选取符合其需求的社会舆情案例总计213件，对事件表述内容借助于ICTCLAS[20]进行分词处理，保留了具有实际意义的名词和动词，辅以人工干预的措施，按照词频选取代表性的词作为各类社会舆情案例的主题标签。对于社会舆情案例的分类，中国传媒大学互联网信息研究院发布的《中国开始进入互联网"新常态"——2014中国网络舆论生态环境研究报告》[21]，将其分为反腐倡廉、司法执法、国防外交、社会民生、时事政治、教育教学、事故灾害、交通运输、企业财经和文化建设10大类，人民网舆情工作室所建案例库将其按领域分为经济生活、社会民生、公共管理、司法事件、文化科教、群体事件、港台国际、地方形象和企业形象9大类[22]，并综合宋余超等[23]提出的分类，将用于确定社会舆情信息采集与组织的特征标签分为：政府管理、社会民生、国防外交、教育教学、事故灾害、企业财经和文化建设7大类。所选取的案例在各类别的分布如下：政府管理类78个、社会民生类51个、国防外交类20个、教育教学类12个、事故灾害类32个、企业财经类13个和文化建设类7个，部分案例如表2所示。

社会舆情案例的标签如表3所示，在具体采集与组织时可以选已经定义好的标签，也可以自主添加标签，政府决策者也可根据自己的需要进行修改和完善。

表 2 用于确定社会舆情信息组织标签的部分案例

社会舆情事件分类	部分案例
政府管理	季建业被查事件、国家药监局局长郑筱萸案、令计划事件、微博开房局长、复旦投毒案、南京彭宇案、钱云会事件、夏俊峰事件、李昌奎案
国防外交	乌克兰政局剧变、中国南海争端事件、习近平马英九会面
社会民生	H7N9 禽流感事件、凤凰古城收费事件、抢盐事件、南航急救门事件、养老金并轨改革、三鹿毒奶粉事件、天价过路费
教育教学	天价幼儿园、南科大退学事件、本科生高考阅卷、河南高考替考事件、人大教授与学生断绝师生关系
事故灾害	雅安芦山强震、马航航班失联、上海元旦夜踩踏事件、“东方之星”长江沉船事故、汶川地震、7.23 动车事件、天津“8.12”爆炸事件
企业财经	家乐福事件、贝因美汞超标事件、小肥羊假羊肉风波
文化建设	国足亚洲杯之旅、嫦娥一号卫星奔月工程、上海世博会、南京梧桐树、第二届世界互联网大会、屠呦呦获得诺贝尔奖

表 3 社会舆情案例组织的部分标签

社会舆情事件分类	组织标签
政府管理	违纪、涉嫌、受贿、滥用职权、通报、走私、嫖宿、行贿、不雅、宣判、刑拘、再审、复查、司法公正、冤案、疑案、重审
国防外交	中方、交会、会面、中国政府、国情、两国、国家安全、海峡两岸、美国
社会民生	病毒、养老、中毒、添加剂、造假、超标、感染、医疗、致癌、急救
教育教学	替考、教改、学术不端、师生关系、外教、学费、退学、贫困生、作弊
事故灾害	地震、强震、火灾、洪灾、泥石流、大雪、相撞、坍塌、辐射、泄漏
企业财经	造假、价格欺诈、不合格、收购、山寨、低品、违规、讨薪
文化建设	诺贝尔、文化保护、历史遗迹、亚洲杯、黄赌毒、赛事、科研成果、科技

2.2 本体应用于社会舆情信息语义组织与支撑政府决策的意义

本体(Ontology)的出现有效地解决了信息组织中的语义冲突方面的问题，其概念最早来源于哲学领域，指研究存在的本质的哲学问题，目前被广泛接受和使用的定义是 1993 年 Gruber[24]提出的“本体是共享的概念模型的形式化规范说明”。本体描述的是某个特定领域的一种形式化、共享的概念化模型，它通过对概念的严格定义和概念之间的关系来确定概念的精确含义，提供对某个领域一种共享的、通用的理解，帮助人和计算机进行交流，其支持语义级的交换[25]。目前不同的领域积极开发自己领域的本体，如生物医药本体、金融本体、法律知识本体、电子政务本体、新闻本体、旅游本体、生物基因本体等，现已有 4 种国内外主要的通用本体库 WordNet、DBpedia、Cyc、HowNet 和两个比较成功的专业领域本体库生物医学和企业领域本体库[26]，然而社会舆情本体的研究鲜有学者涉猎。社会舆情信息组织是一个复杂的过程，目前此方面的研究大多停留在信息的结构化聚合层面，未建立社会舆情信息的多维度关联体系。随着大数据时代的来临，研究者的着眼点已不再局限于揭示信息之间的因果关系，而是更关注相关关系，通过应用相关关系，可以比以前更容易、更快捷、更清楚地分析事物。

在政府公共决策制定与实施过程中，技术发展带来的有关民情、民意、民智等社会舆情信息与知识的传播日益复杂化，然而在混乱纷杂的数据中寻找有用的信息和知识也成为政府决策者必须面对的挑战。基于本体的社会舆情信息组织研究对于政府部门及时、高效地获取有序化、集成化和知识化的社会舆情信息，并采取有效措施干预与引导社会舆情的走向具有实践指导意义，是社会舆情研究对于政府决策突破困境的有力支持。

单个社会舆情分析案例都是宝贵的知识财富，通过对社会舆情个案进行不断的存储和积累，形成不同类型社会舆情事件的历史案例数据库，通过对不同的社会舆情个案进行关联和纵向分析，能够提炼出相关社会舆情的规律和特点，可以增强对同类型舆情事件的认知和理解，有助于政府在面临新的社会舆情事件时，进行更加精确的推断和预测，洞悉其未来的演化趋势。有了社会舆情本体库的支撑，政府部门在决策时能够减少因数据缺乏而造成的分析偏差，也可以避免个人意志带来的主观性、片面性和局限性决策，从而优化和完善公共事务决策流程，使数据驱动的社会决策与科学治理常态化，为社会有序运行提供可靠的保证。

2.3 基于本体的社会舆情信息语义组织思路

本研究选取社会舆情为研究对象，主要对社会舆

情的事件所属主题、参与主体、发生地点、发生时间、网民情感倾向等方面的概念及概念间的关系进行描述。借助本体所建立的关联，将社会舆情信息的多维语义特征内容进行整合，由于设置的案例知识库的字段更加细粒度，更容易为新增事件提供相关案例的参考。构建社会舆情本体的目标是要形成对于社会舆情信息组织结构的共同认可、理解并分析舆情领域的知识，为其他领域本体的构建提供有章可循的方法和技术路线，同时为进一步建立舆情领域语义 Web 奠定坚实的基础。

针对从新闻报道、微博、博客、论坛等各种舆情信息源采集的社会舆情信息，分析主要涉及的概念，构建社会舆情信息本体的概念模型。社会舆情信息语义组织的框架可从外部特征和内容特征两条主线进行设定，社会舆情信息的外部特征如标题、时间、来源等，可通过简单的预处理、消歧、过滤等操作获取，而内容特征如人物、地点、机构、媒体、事件、主题、情感等，这些知识单元抽取的步骤主要包括中文分词、词性标注、命名实体识别、去停用词、情感提取。从外部特征和内容特征两个方面展开舆情特征的抽取工作，并将抽取的知识单元基于构建的社会舆情信息本体揭示其内在的多维度关联，最后存储到数据库中。

3　社会舆情信息语义组织框架的建立

3.1　社会舆情信息本体构建方法

目前已有几种经典的本体构建方法，如 IDEF、骨架法、TOVE 法、METHONTOLOGY 法、七步法等[27-28]，但还未形成统一的标准。总的来说，主要有两种本体构建思路：一种是利用本体技术直接构建本体模型；另一种是借鉴或者复用已有的成熟本体进行构建。本研究的本体构建将按照七步法的步骤，借鉴已有的成熟本体，在分析现有事件模型的基础上，构建基于本体的社会舆情信息模型(Public Opinion Information Ontology Model, POIOM)。

虽然尚未有包含各类社会舆情信息的成熟本体，但事件是社会舆情信息的依托，已有一些成熟的事件本体模型，如 ABC[16]、PROTON[29]、Event-Model-F[30]等，相比较而言，ABC 本体具有轻量级、通用性好且概念清晰等优势，能够对一般事件过程进行良好的表达[31]，本研究以 ABC 本体作为借鉴对象；此外，也有一些研究者着眼于具体领域的舆情本体的研究，如旅游危机事件网络舆情本体[14]、食品安全网络舆情本体[15]等，这都是本研究构建 POIOM 的参考。

3.2　社会舆情信息本体 POIOM 建模

POIOM 建模的目标是帮助舆情事件参与主体，特别是政府机构在舆情发展变化的关键节点及时有效地采取措施、解决问题、平息舆论风波。为了更有效地表达社会舆情案例的发展过程和知识结构，本文在 ABC 本体的基础上进行精简和扩展，增加个人、组织等主体概念和情感、主题等概念，具体表述如下：

(1) 个人(Person)和组织(Organization)都是概念 Agent 的扩展子类，个人和组织作为参与主体共同作用于社会舆情事件，影响其发展与演变。为了更好地体现参与主体中政府机构的应对效应，将组织类进一步分为 Government、Media、Social 三个子类，分别表示政府、媒体、社会组织这三类实体，为了避免概念的混淆，这里的媒体不包含在社会组织中，社会组织界为除政府、媒体之外的各类企业、单位、群体、团队等。

(2) 用情感倾向(Sentiment)来表达某个事件发生后的情感状态(Situation)，具有时间依赖性。

(3) 用舆情主题(Topic)来表示社会民众在很长时间内普遍关注的中介性社会事项，利于政府决策者对舆情整体的把握。

在构建 POIOM 模型时，将社会舆情领域信息分成了层：事件层、过程层、行动层、决策层，如图 1 所示。其中，描述性的知识定义在事件层，指定社会舆情事件相互之间的关系；过程性知识主要定义在过程层，描述了社会舆情事件涉及到的一系列情感状态；行动层主要定义政府回应社会舆情事件所采取的行动知识；决策层主要定义社会舆情事件应对方案、应急预案等知识，层之间的关系包括参与、响应、参考等。

3.3　社会舆情信息的语义关联特征抽取

(1) POIOM 中的概念及其界定

本研究定义的社会舆情信息本体 POIOM 的概念如下所示：

| Thing | Happening | Time | Place | File | Resource | Agent | Topic | Event | Sentiment | Action | Case | Person | Organization | Media | Social | Government | News | Weibo | Government_management | Social_living | Defence_diplomacy | Education | Corporate_finance | Accident_disaster | Cultural_construction |

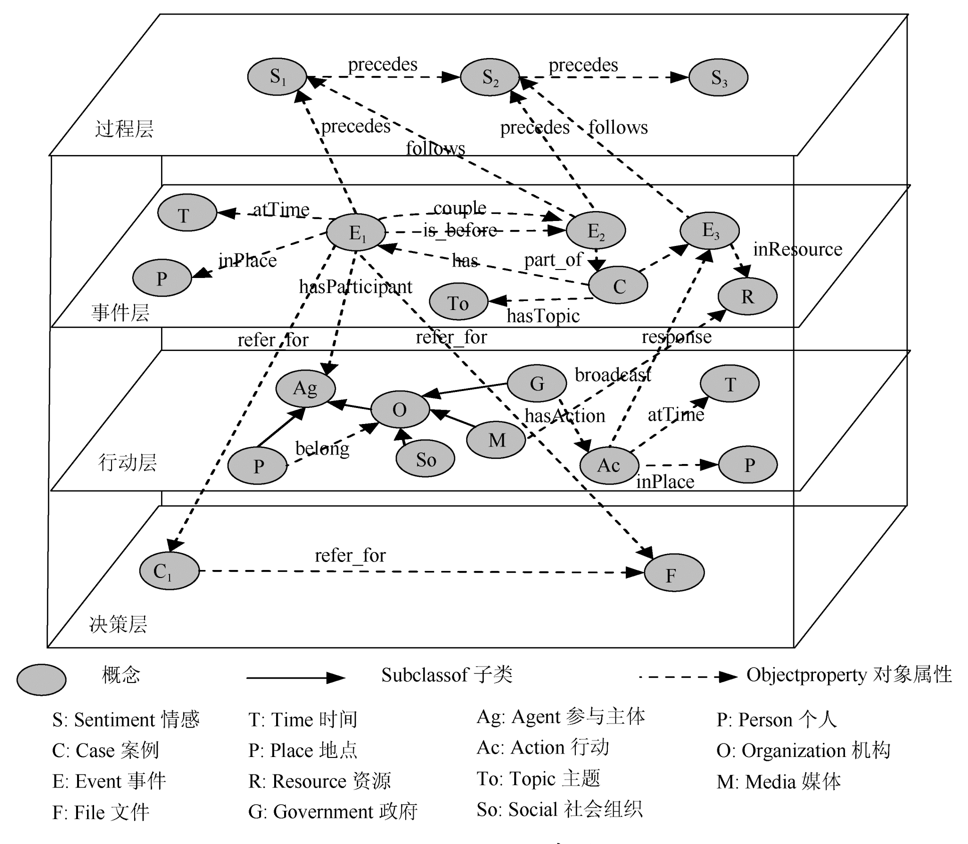


图 1　POIOM 建模

社会舆情信息本体主要包含 1 个顶级类 Thing、7 个一级类、15 个二级类和 3 个三级类，其中 Thing 是万物之源，其余涉及的概念和类具体如下：

①Happening 是所有舆情事件的父类，Event、Sentiment、Action 和 Case 是其 4 个子类，各子类的具体含义如下：Event 表示动态事件的概念，多个 Event 共同组成一个具体的案例；Sentiment 描述事件发生之前及后续状态，以事件发展过程中的情感表现，主要是针对微博发文、微博评论及新闻跟帖的情感倾向分析结果的体现；Action 也指发生的事件，但它有明确主体，侧重描述参与主体采取的针对 Event 事件的行动措施；Case 是指具有关联关系的事件集，它能够描述由一个关系事件引起的，并被相关子事件推动发展的一个完整的案例。

②Agent 是 Person、Organization 的父类，用于表示事件中的 Who/Whom 要素，即施事者和受事者。其中 Organization 下又包括 Government、Media、Social 三个子类，分别表示政府、媒体、社会组织这三类实体。

③Topic 表示舆情主题，包含 Defence_diplomacy、Government_management、Education、Social_living、Corporate_finance、Accident_disaster、Cultural_construction 这 7 个子类，分别表示国防外交、政府管理、教育教学、社会民生、企业财经、事故灾难、文化建设这几类舆情类型，具体划分标准参考宋余超等[23]对舆情主题的划分，每类舆情事件都有对应的舆情主题，当有新主题时可以自定义添加补充，如表 4 所示：

表 4　各类舆情主题部分示例

社会舆情事件分类	主题
政府管理	反腐倡廉、司法执法、官员形象、信息透明、办事质量
国防外交	港台关系、涉外关系、民族宗教、恐怖袭击
社会民生	食品安全、医疗、就业、环境卫生、住房拆迁、社会保障
教育教学	教育公平、教育改革、学术诚信、师生关系
事故灾害	自然灾害、交通事故、安全生产
企业财经	企业形象、雇佣矛盾、标准生产、物价浮动
文化建设	体育赛事、庆祝活动、娱乐整顿、明星黄赌毒

④Time 表示事件涉及的时间，包括案例的开始时间和结束时间、事件发生的时间、行动发生的时间等，它可以是十分明确、具体的时间指示词，如“2015 年 12 月 12 日”，也可以是类似“昨天、两天前”这样的相对时间表述词。

⑤Place 表示事件涉及的地点，一般由具体的地名展示。

⑥File 表示已有的应急预案、国家的政策、标准等，可供具体的社会舆情事件案例参考。

⑦Resource 是事件的载体，指向具体的资源数据，包括 News、Weibo 两个子类，分别表示新闻资源和微博资源。

(2) POIOM 中概念的数据属性

对于 POIOM 中的每个概念都有数据属性，其中子类继承父类的全部属性，因此子类的某些与父类相同的属性不再赘述，具体如表 5 所示：

表 5 社会舆情信息本体数据属性

类名	父类属性	一级子类属性	二级子类属性
参与主体(Agent)	中文名称、英语名称		
个人(Person)		角色、别称、籍贯、性别	
组织(Organization)		负责人、简介	
政府(Government)			级别、职责
媒体(Media)			类型、成立时间、级别
社会组织(Social)			类型、规模、经营范围、所属行业
舆情事件(Happening)	名称、类型、描述、造成的损失		
案例(Case)		新闻报道量、微博发文量、评论量、参与量	
情感(Sentiment)		情感级别、负面情感比例值	
地域(Place)	地域名、所属省区、经度、纬度		
时间(Time)	年、月、日、时、分		
资源(Resource)	资源类型、采集关键词、网址、发布者、发布时间、发布内容		
新闻(News)		标题、来源、跟帖数、参与数	
微博(Weibo)		转发数、评论数、点赞数、来源	
文件(File)	文件名、文件内容、出台时间、出台机构		

3.4 建立社会舆情信息语义关联

事件及其涉及的实体可能会与其他事件(或实体)相关，社会舆情事件与事件之间也存在此类关联，这些关系可以是时间上并发、空间上并现，也可以是时序关系、相似关系、耦合关系、因果关系等，通过语义关联的获取，可以得到相关事件的集合。

本研究总结了 POIOM 本体中涉及的概念间关系，如表 6 所示，其中省略了子类与父类之间的 subClassOf 关系。

表 6 POIOM 的语义关系及其示例

概念间关系	概念名		示例	
相似关系(resemble)	Case	Case	天津港特别重大火灾爆炸事件	福建漳州 PX 项目爆炸事件
时间关系(is_before、is_after)	Event	Event	爆炸	污染
并发关系(synchronize)	Event	Event	骚乱	踩踏
因果关系(cause、causeby)	Event	Event	台风	暴雨
顺序关系(precede、follow)	Event	Situation	救援	1 级
参考关系(refer_for)	Case	File	台湾地沟油事件	国家食品安全事故应急预案

(续表)

概念间关系	概念名		示例	
部分–整体关系(part_of)	Event	Case	起火	杭州公交纵火事件
包含关系(has)	Case	Event	杭州公交纵火事件	起火
响应关系(response)	Action	Event	立案调查	起火
参与关系(hasParticipant)	Case Event	Agent	郭美美事件 道歉	郭美美
在……时间舆情开始(startTime)	Case	Time	南京急救门事件	2015 年 11 月 22 日
在……时间发生(happenTime)	Event Action Case	Time	微博爆料 立案调查 南航急救门事件	2015 年 11 月 22 日 2015 年 11 月 26 日 2015 年 11 月 09 日
位于关系(inPlace)	Event Agent Action Case	Place	突发 北京市卫计委 核实 南航急救门事件	北京
来源关系(inResource)	Event	Resource	致歉	《台湾“行政院长”因问题油事件致歉》
归属关系(belong)	Person	Organization	马云	阿里巴巴(中国)网络技术有限公司
报道关系(broadcast)	Media	News	中国新闻网	“北京红会紧急救援中心向南航急救事件当事人致歉”
采取关系(hasAction)	Government	Action	北京市卫计委	立案
主题关系(hasTopic)	Case	Topic	南航急救门事件	医疗

4　社会舆情信息语义组织的实现与管理

4.1　本体构建与实现结果

通过上文的研究，可将社会舆情信息本体 POIOM 形式化表示为: DomO={C, R_H, R_P, I, A}，其中 C 代表概念; R_H 指概念之间或实例之间的层次关系; R_P 指概念之间或实例之间的语义关系，其将不同的概念和不同的实例联系起来; I 指实例，代表某个类的具体实体; A 指公理，代表永真断言。

采用 Protégé [32]实现了社会舆情信息本体的构建，实现本体类、属性、关系的定义，并且为每个类创建对应的实例，填充各个实例的属性值。Protégé 4.3 还提供了本体的可视化展示，加载 Graphviz 2.28 插件，选择 Protégé 4.3 中的 OWLViz 选项，将所构建的社会舆情信息本体可视化地展示出来，具体实现的部分结果如图 2 和图 3 所示。

4.2　本体管理模块设计

社会舆情信息采集及组织系统以为政府决策提供结构化、有序、具有语义关联的社会舆情信息为系统设计出发点，相当于信息获取与组织的工具，应包括 4 大功能:

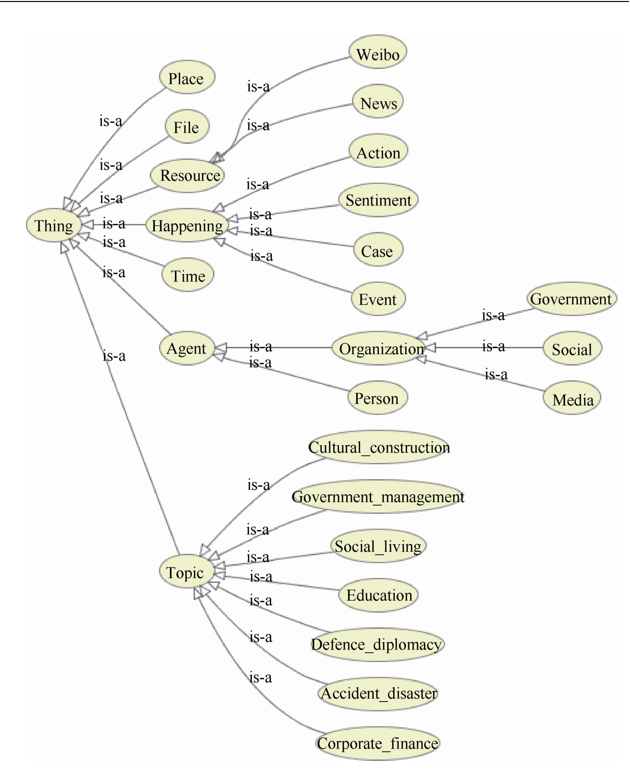


图 2　社会舆情本体的树图

(1) 信息采集: 利用网络爬虫技术针对新闻媒体、微博等舆情信息源进行采集，可选择具体社会舆情事

件类型领域，也可以针对某个具体的采集关键词实施采集；

(2) 语义特征抽取：通过中文分词、命名实体识别、去停用词、情感倾向性分析等代码的调用，抽取出人名、组织机构名、媒体名、事件、行动等信息，并进行存储；

(3) 本体管理：可实现对所建本体的管理，不需重新利用本体编辑软件重新写回修改，可不断根据政府决策者的需求对所建的社会舆情信息本体 POIOM 进行修改和完善；

(4) 信息检索：提供社会舆情信息检索功能，可对具体实例信息进行检索，能够展示搜索结果及其关联信息，也可查看到该实例的具体信息。

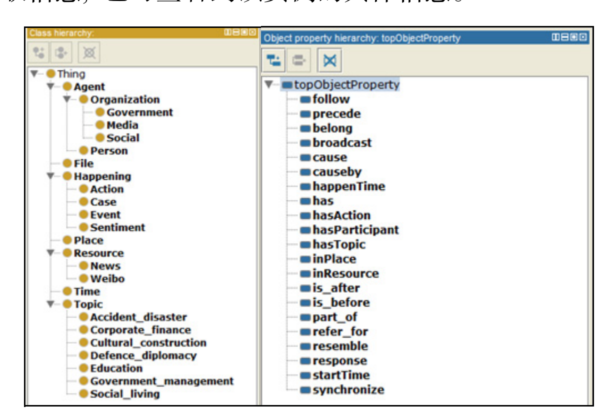


图 3 社会舆情本体的部分实现结果

系统的核心建立在 POIOM 本体的基础上，对本体的管理是系统的主要功能模块。本体管理模块主要包括两部分的内容：类管理与属性管理，具体界面如图 4 和图 5 所示，可进行添加属性值、修改属性值、删除属性值、添加关系、修改关系和删除关系等本体管理操作。



图 4 “南航急救门事件”案例类管理



图 5 “南航急救门事件”案例实例管理

5 结 语

社会舆情信息是最有价值的社会信息资源，对其进行语义组织有助于进行深层次的社会舆情信息资源开发，为社会提供更好的信息和知识服务。本研究以面向政府决策需求为出发点，基于语义关联的理论构建了社会舆情信息本体模型，并对其实现步骤等进行探讨。然而，社会舆情信息语义组织工作是一项复杂的系统工程，今后需要在进一步梳理与细致界定政府决策需求的基础上，对社会舆情本体进化以及数据关联关系利用等方面继续开展深入研究。

参考文献：

[1] 中国图书馆学会. 2011-2012 年图书馆学学科发展报告[M]. 北京：中国科学技术出版社, 2012: 1-5. (Library Society of China. Report on Advances in Library Science [M]. Beijing: Science and Technology China Press, 2012: 1-5.)

[2] Sure Y, Akkermans H, Broekstra J, et al. On-To-Knowledge: Semantic Web Enabled Knowledge Management [C]. In: Proceedings of Web Intelligence.Springer-Verlag, 2003: 277-300.

[3] Staab S, Maedche A, Handschuh S. An Annotation Framework for the Semantic Web [C]. In: Proceedings of the 1st Workshop on Multimedia Annotation, Tokyo, Japan. 2001: 30-31.

[4] 方平. 试论一体化医学语言系统(UMLS)超级叙词表的特点[J]. 图书情报工作, 1998, 42(10): 26-29, 41. (Fang Ping. The Characteristics of the UMLS Metathesaurus [J]. Library and Information Service, 1998, 42(10): 26-29, 41.)

[5] 李萍萍, 李书宾. 基于语义 Web 的知识组织方法与应用[J]. 中华医学图书情报杂志, 2009, 18(4): 8-11. (Li Pingping, Li Shubin. Semantic Web-based Knowledge Organization

Methods and Their Application [J]. Chinese Journal of Medical Library and Information Science, 2009, 18(4): 8-11.)

[6] 苏晓路，钱平，颜蕴，等．农业科技信息智能检索系统中的知识组织[J]．现代图书情报技术，2005(12): 34-38. (Su Xiaolu, Qian Ping, Yan Yun, et al. The Knowledge Organization in the Agricultural Intelligent Retrieval System [J]. New Technology of Library and Information Service, 2005(12): 34-38.)

[7] 董慧，杜文华．基于本体和多代理的数字图书馆信息检索模型[J]．中国图书馆学报，2004, 30(2): 63-65. (Dong Hui, Du Wenhua. A Modal of Digital Library Information Retrieval Based on Ontology and Multiple Agents [J]. Journal of Library Science in China, 2004, 30(2): 63-65.)

[8] 董慧，王菲，姜赢，等．基于数字图书馆的本体应用环境研究[J]．中国图书馆学报，2009, 35(5): 52-58. (Dong Hui, Wang Fei, Jiang Ying, et al. A Study of the Ontology Application Environment Based on the Digital Library[J]. Journal of Library Science in China, 2009, 35(5): 52-58.)

[9] 申欣，田爱奎，朱瑞瑾．基于语义网的信息检索系统的研究与实现[J]．山东理工大学学报：自然科学版，2011, 25(1): 92-97. (Shen Xin, Tian Aikui, Zhu Ruijin. Research and Implementation of Information Retrieval System Based on Semantic Web [J]. Journal of Shandong University of Technology: Natural Science Edition, 2011, 25(1): 92-97.)

[10] Kinsella S, Breslin J G, Passant A, et al. Applications of Semantic Web Methodologies and Techniques to Social Networks and Social Websites [A]. // Reasoning Web [M]. Springer Berlin Heidelberg, 2008: 171-199.

[11] Sleeman J, Finin T. Computing FOAF Co-reference Relations with Rules and Machine Learning [C]. In: Proceedings of the 3rd International Workshop on Social Data on Web, Shanghai, China.2010: 595-608.

[12] 王兰成，徐震．主题舆情分析中的语义 Web 文本分类研究[J]．情报学报，2012, 31(4): 340-344. (Wang Lancheng, Xu Zhen. Research on Semantic Web Text Categorization in Subject Public Opinion Analysis [J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(4): 340-344.)

[13] 王兰成．网络舆情情报语义集成分析法的理论与实践探析[J]．情报理论与实践，2013, 36(2): 48-50. (Wang Lancheng. Analysis on Theory and Practice of Intelligence Semantic Association Integrates Analysis [J]. Information Studies: Theory & Application, 2013, 36(2): 48-50.)

[14] 付业勤，郑向敏，雷春，等．旅游危机事件网络舆情的本体分析[J]．重庆理工大学学报：社会科学，2015, 29(2): 47-53. (Fu Yeqin, Zheng Xiangmin, Lei Chun, et al. Analysis on Network Opinion Ontology of Tourism Crisis [J]. Journal of Chongqing University of Technology: Social Science, 2015, 29(2): 47-53.)

[15] 李宏伟，林萍，洪小娟．食品安全网络舆情本体学习研究[J]．南京邮电大学学报：社会科学版，2013, 15(4): 72-77. (Li Hongwei, Lin Ping, Hong Xiaojuan. Research on Ontology Learning of Internet Public Sentiment Concerning Food Safety [J]. Journal of Nanjing University of Posts and Telecommunications: Social Science, 2013, 15(4): 72-77.)

[16] Lagoze C, Hunter J. The ABC Ontology and Model [J]. Journal of Digital Information, 2002, 2(2): 1-18.

[17] 方兴东，熊剑．网络舆情蓝皮书[M]．北京：电子工业出版社, 2015: 34-60. (Fang Xingdong, Xiong Jian. The Blue Book of Net-mediated Public Sentiment [M]. Beijing: Publishing House of Electronics Industry, 2015: 34-60.)

[18] 谢耘耕．中国社会舆情与危机管理报告[M]．北京：社会科学文献出版社，2014: 1-10. (Xie Yungeng. The Report on Chinese Social Opinion and Crisis Management [M]. Beijing: Social Sciences Literature Press, 2014: 1-10.)

[19] 杜骏飞，邹鸿强．危如朝露：2010～2011 中国网络舆情报告[M]．杭州：浙江大学出版社, 2011: 7-10. (Du Junfei, Zou Hongqiang. Report on Public Opinion of China Internet 2010-2011 [M]. Hangzhou: Zhejiang University Press, 2011: 7-10.)

[20] ICTCLAS 汉语分词系统[EB/OL]. [2015-11-25]. http://www.ictclas.org. (ICTCLAS, Institute of Computing Technology, Chinese Lexical Analysis System [EB/OL]. [2015-11-25]. http://www.ictclas.org.)

[21] 张树庭，李未柠，孔清溪．中国开始进入互联网“新常态”——2014 中国网络舆论生态环境研究报告[J]．现代传播，2015, 37(3): 1-9. (Zhang Shuting, Li Weining, Kong Qingxi. China Began to Enter the Internet New Normal——The 2014 Report on the Ecological Environment of the Network Public Opinion in China [J]. Modern Communication, 2015, 37(3): 1-9.)

[22] 人民网舆情工作室案例库[EB/OL]. [2015-12-17]. http://yq.people.com.cn/CaseLib.htm. (Cases in Public Opinion Monitoring Office of People.cn [EB/OL]. [2015-12-17]. http://yq.people.com.cn/CaseLib.htm.)

[23] 宋余超，陈福集．面向主题分类的网络舆情博弈形成机制研究——基于三主体研究[J]．情报杂志，2015, 34(8): 122-126, 121. (Song Yuchao, Chen Fuji. The Formation Mechanism of Online Public Opinion Based on Topic Classification and Game Theory: A Study Based on the Three

Subjects Involved [J]. Journal of Intelligence, 2015, 34(8): 122-126, 121.)

[24] Gruber T R.A Translation Approach to Portable Ontologies Specifications [J]. Knowledge Acquisition, 1993, 5(2): 199-220.

[25] 白同强, 刘磊. 语义 Web 的研究与展望[J]. 吉林大学学报: 信息科学版, 2004, 22(2): 154-159. (Bai Tongqiang, Liu Lei. Research and Prospect in Semantic Web [J]. Journal of Jilin University: Information Science Edition, 2004, 22(2): 154-159.)

[26] 白如江, 于晓繁, 王效岳. 国内外主要本体库比较分析研究[J]. 现代图书情报技术, 2011(1): 3-13. (Bai Rujiang, Yu Xiaofan, Wang Xiaoyue. The Comparative Analysis of Major Domestic and Foreign Ontology Library [J]. New Technology of Library and Information Service, 2011(1): 3-13.)

[27] 杜文华. 本体构建方法比较研究[J]. 情报杂志, 2005, 24(10): 24-25. (Du Wenhua. The Comparative Analysis of Approaches in Constructing Ontology [J]. Journal of Information, 2005, 24(10): 24-25.)

[28] 刘宇松. 本体构建方法和开发工具研究[J]. 现代情报, 2009, 29(9): 17-24. (Liu Yusong. Research of Approaches and Development Tools in Constructing Ontology [J]. Journal of Modern Information, 2009, 29(9): 17-24.)

[29] Kiryakov A, Popov B, Ognyanoff D, et al. Semantic Annotation, Indexing, and Retrieval [M]. The Semantic Web - ISWC 2003. Springer Berlin Heidelberg, 2003: 484-499.

[30] Scherp A, Franz T, Saathoff C, et al. F--A Model of Events Based on the Foundational Ontology dolce+DnS Ultralight [C]. In: Proceedings of the 5th International Conference on Knowledge Capture. ACM, 2009: 137-144.

[31] 王文俊, 杨鹏, 董存祥. 应急案例本体模型的研究及应用[J]. 计算机应用, 2009, 29(5): 1437-1440, 1445. (Wang Wenjun, Yang Peng, Dong Cunxiang. Study and Application of Emergency Case Ontology Model [J]. Journal of Computer Applications, 2009, 29(5): 1437-1440, 1445.)

[32] Protégé [EB/OL]. [2013-11-30]. http://protege.stanford.edu/.

作者贡献声明:

王曰芬: 提出研究思路, 拟定研究方案, 论文修改定稿;
邢梦婷: 采集数据, 构建模型, 撰写论文初稿。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: 274686493@qq.com。

[1] 邢梦婷, 王曰芬. 访谈提纲.doc.关于"政府舆情信息需求情况"的访谈提纲.

[2] 邢梦婷, 王曰芬. 案例列表.doc. 2007 年–2014 年政府关注的社会舆情案例.

[3] 邢梦婷, 王曰芬. 语义关联特征抽取方法.doc. 语义关联特征抽取方法.

[4] 邢梦婷, 王曰芬. 南航急救门事件发展过程.doc. 南航急救门事件发展过程.

收稿日期: 2016-06-12
收修改稿日期: 2016-07-17

Semantic Analysis of Social Public Opinion Information for Government Decision Making

Wang Yuefen[1,2]　Xing Mengting[1]
[1](School of Economics and Management, Nanjing University of Science &Technology, Nanjing 210094, China)
[2](Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094, China)

Abstract: [**Objective**] This paper identifies multi-dimensional associations among social public opinion data with the help of semantic organization analysis technology. It also provides important support for data mining and decision-making tasks in the future. [**Methods**] We first studied the governmental decision-making needs to establish a semantic organization framework for the public opinion information. Second, we analyzed the characteristics of these information and then adopted the "seven-step-method" as well as the ABC event Ontology model to create a new one. Finally, we established the semantics connection among the public opinion information elements. [**Results**] We proposed a new POIOM model and management module for the public opinion information. [**Limitations**] This paper used interviews and statistical analysis to assess the information needs. More participants should be included in future studies. [**Conclusions**] The POIOM model could effectively describe the social public opinion information. It can also reveal the association among different features of the same case, as well as the logical and internal relationship among multiple public opinion events. This study shows that the semantic organization method is feasible and effective to process social public opinion information.

Keywords: Government decision demands　Public opinion　Ontology　Semantics organization

ProQuest 与 USC Shoah 合作提供种族灭绝目击者视频证词的访问

ProQuest 已经与 USC Shoah 视觉历史和教育研究所合作发布了视觉历史档案(Visual History Archive)的流文件版本，提供了种族灭绝幸存者和目击者共 53 000 份视频证词。

此流文件版本包括新的 ProQuest 搜索功能，使用户能够找到特定的关键词和相关的 ProQuest 内容——数十亿可供搜索的内容，包括学位论文、新闻、杂志文章、学术期刊文章和电子书。对于图书馆来说，访问视觉历史档案不再需要专用的互联网连结和缓存服务器，大大降低了成本，消除了下载延迟。

视觉历史档案包括超过 112 000 小时的犹太人大屠杀和其他种族灭绝行动，包括亚美尼亚大屠杀，卢旺达种族大屠杀和南京大屠杀的幸存者和目击者的视频证词。视频访谈遍及 63 个国家，以大约 40 种语言进行，探索种族灭绝行动之前、期间和之后的生活情况。整个档案的视频量非常大，如果一天 24 小时不停地播放，需要 13 年才能播完。但是，通过 62 000 个人工索引的搜索词，研究人员可以优化搜索结果并精确到分钟。

ProQuest 与 USC Shoah 的合作旨在拓宽视觉历史档案的曝光度。这一合作补充了 ProQuest 现有的索引方法，帮助用户检索感兴趣的特定点的视频内容。

此外，视觉历史档案还在逐年增加中。仅在 2016 年，就将增加 1 000 份来自柬埔寨和危地马拉种族灭绝以及犹太人大屠杀的视频证词。

2017 年年末，ProQuest 将在整个 ProQuest 内容中，包括大量的历史资源库，如历史报纸、期刊、杂志、政府记录和其他原始资料库中，提供对这些视觉历史档案视频证词的检索。例如，订阅了 ProQuest 历史犹太报纸的图书馆，可以添加视觉历史档案到他们的馆藏中，从而为其服务的学生和研究人员提供更丰富的体验。

(编译自: http://www.proquest.com/about/news/2016/Landmark-Partnership-Improves-Access-to-Eyewitness-Accounts.html)

(本刊讯)